THE INTERPLAY AND INTERACTION BETWEEN PROSODY AND SYNTAX: EVIDENCE FROM MANDARIN CHINESE

Chiu-yu TSENG and Da-de CHEN
Institute of Linguistics (Preparatory Office)
Academia Sinica
Taipei, China 115-21
e-mail: cytling@sinica.edu.tw

ABSTRACT
This paper aims to (1.) quantify possible correlation between syntactic structure and prosodic manifestation in Mandarin Chinese, (2.) explore to what extent such correlation could be predicted by syntactic structures and what may go beyond these correlations, and (3.) increase more operational characteristics to Mandarin prosodic structures.

INTRODUCTION
A major portion of the previous works by our group on a Mandarin Chinese speech database collected in Taiwan have been concentrated on developing tools to process the collected speech data [1]. Since we attempted to process the speech data from the physical perspective by employing a phonetic oriented approach, the development of these tools depended largely on how we characterized the major phonetic features derived. We also aimed to characterize the major phonetic as well as prosodic properties of Mandarin Chinese [2]. Note that these tasks were performed under the spirit of ToBI [3] so that they could be compared with similar tools developed for other languages. The issues we have dealt with over time ranged from the segmental layers to the prosodic layers, the latter being our more recent focus. Our initial effort to characterize labeling system at the prosodic layer included designed a labeling system for breaks/pauses in continuous speech. The ToBI group has since acknowledged and adopted our break labeling system for Mandarin at the prosodic layer.

We noted that the organization of prosody, as with the organization of segmental phenomena, can be seen from different perspectives. Analyzing the manifestation appeared to be one from the most logical aspect if we approach the speech data from the phonetic point of view. However, though our previous approach focused on the phonetic aspects of continuous speech, that is, our analyses and experiments emphasized the physical as well as perceptual properties of speech sounds only, we have not overlooked the fact that speech output is in fact the ultimate derivation of many other parallel linguistic levels of information interacting and interplaying at the same time. In other words, any particular speech event can also be seen in that light. Take pauses/breaks for example, labeling them through a listening test means working with the physical data only and largely trying to characterize these phenomena from the perceptual point of view. In this particular chosen frame of work, arriving at perceptual consistency within and among transcribers was crucial since their labeling results would serve as the basis of the development of labeling tools. But if we take a step away from the physical data only, a reasonable question would be: Can we look beyond the lower linguistic level of phonetic information to other higher linguistic levels for more information that we know are significant, or even, for possible governing factors that to a large extent contribute to the occurrences of these breaks/pauses? In short, what could be the major governing factors of prosodic organization [4,5,6]?

The syntactic structures involved are definitely one of the most linguistically feasible factors to consider [7,8], especially with respect to syntactic boundaries. In this paper, we will report our attempt to investigate whether such interplay can be derived through correlation measurements between perceived breaks/pauses and syntactic boundaries. We will look into the prosodic manifestation and syntactic boundaries through a correlation analysis between manual labeling results with the syntactic parsing analyses.

EXPERIMENTS
Based on the assumptions that (1.) the position where breaks/pauses occur in running speech may correspond to most but not necessarily all major syntactic boundaries and (2.) breaks/pauses should not occur across syntactic boundaries, we carried out two independent kinds of labeling and later mapped the results. 1107 prosody oriented Mandarin utterances from our speech database were used. These utterances are continuous read speech by one female speaker in her late 20's; they range from 7 to 140 syllables/characters in duration. Of these 1107 utterances, 804 are declarative utterances and 303 are exclamatory utterances in structure. The two independent kinds of labeling are: (1.) the perceived breaks that occurred in the speech data of these utterances and (2.) derived syntactic boundaries through text analysis software. To avoid inter-transcriber inconsistencies, only one trained transcriber’s labeling results were used. The transcriber labeled the breaks through listening tasks using a ToBI-spirited labeling system developed for Mandarin by our group [2]. The 5-increment 6-step break labeling system (B0 to B5) characterizes breaks as follows: B0 corresponds to reduced syllabic boundary, B1 normal syllabic boundary, B2 minor phrase boundary, B3 major phrase boundary, B4 breath-group boundary and B5 prosodic group boundary. The labeling results used represent perceived breaks by one human listener. Figure 1 shows an example of a labeled utterance for breaks perceived.
For each kind of break labeling, we then studied the frequencies of its correspondence to the 11 possible brackets. The following example shows another analysis of a much longer utterance without the waveforms.

We then mapped the above two independent labeling results together by utilizing each break label individually. The resulting labeling structure is shown in Table 1.

Table 1 summarizes the results of correlation by utterance types.
corresponding syntactic boundaries in exclamatory utterance

Table 1a shows that in the 804 declarative utterances, most of B2 (minor phrase boundary) correspond to two kinds of brackets, namely, 0 bracket (35.8%) and 1 bracket (31%). The latter would be the most reasonable correspondence, the former may attribute to speaker’s intension and/or some kind of emphasis. Note that the next highest correspondence occurred for 2 brackets (19.6%). B3 (major phrase boundary) corresponded mostly to 2 brackets (29.3%) and 3 brackets (21.6%). Note also that 13.9% of B3 corresponded to 4 brackets. B4 (breath-group boundary) corresponded mostly to 2 brackets (24.7%), 3 brackets (23.6%) and 4 brackets (20.7%) whereas B5 (prosodic group boundary) corresponded mostly to 2 brackets (28%), 3 brackets (22.1%) and 4 brackets (22.5%). Note that 11.5% of B4 corresponded to 5 brackets whereas 14.4% of B5 also corresponded to 5 brackets.

Table 1b shows that in the 303 exclamatory utterances, most of B2 (minor phrase boundary) correspond to two kinds of brackets, namely, 0 bracket (34.8%) and 1 bracket (36.3%). Note that the next highest correspondence occurred for 2 brackets (18%). We also note that B2 behaved somewhat similarly for both declarative and exclamatory utterances. B3 (major phrase boundary) corresponded mostly to 2 brackets (36%) and 3 brackets (22.7%). Note that 15.9% of B3 corresponded to 1 bracket while 12.5% of B3 corresponded to 4 brackets. In other words, B3 is seen to correspond differently to numbers of syntactic bracketing in declarative and exclamatory utterances. B4 (breath-group boundary) corresponded mostly to 2 brackets (36.7%) and 3 brackets (29.1%) and 4 brackets (20.7%) whereas B5 (prosodic group boundary) corresponded mostly to 2 brackets (25.7%) and 3 brackets (22.1%) and 4 brackets (26.1%). Note that 15.5% of B5 corresponded to 1 bracket and 16.9% of B5 corresponded to 4 brackets.

**DISCUSSION**

Though we were able to obtain some kind of correlations, they are by no means high. However, we believe that we should be able to improve the correlation later on. As mentioned above, the CKIP parser is not only a syntactic parser, it also included lexical features and thematic structures. The latter may contribute to the high number of brackets designed. Secondly, 11 layers of bracketing may be too many for our kind of correlation, therefore breaking up the correspondence. One direction would be to collapse the number of brackets into larger syntactic categories in our next step, thereby reducing the steps from 11 to less numbers. The rationale would be that we would be interested in looking into larger syntactic units only, such as noun phrases (NP), verb phrases (VP) and so forth in the purely syntactic sense. In other words, given the amount of manually labeled prosodic data available, we should probably aim at fine-tuning the syntactically parsed results for possible better correlation. Thirdly, the unexpected occurrence of B5 (prosodic group boundary), the longest break/pause, in the middle of a long utterance is worth more detailed investigation. We reason here that breaks could be used together with prominence, another layer of prosodic property currently under investigation. We intend to integrate systematically labeled prominence information in our next project, too.

**CONCLUSION**

In this study, we reported our first attempt to correlate the relationship between syntactic structure and actual prosodic manifestation in continuous speech of Mandarin Chinese, using perceived breaks/pause as our prosodic reference. Independent labeling/tagging results of the two linguistic levels were mapped to test if correlation could be found. We also see the reported attempt as a first step towards a working model for the organization of prosody. The linguistic significance here is quite clear: the interaction between breaks/pauses and syntactic boundaries can be seen through this kind of investigation. A speech database certainly offered a much wider range of evidence that furthered our knowledge in this respect. Our preliminary results showed that mapping was indeed found. However, if the proposed type of correlation were to be utilized for application in speech science research or tools, finer tailoring of the mapping would be necessary. Nevertheless, direct application to speech synthesis would be a highly likely area, where speech recognition could also benefit from the kind of results obtained. We believe that the incorporation of more levels of prosodic information such as prominence and emphasis would be necessary and crucial to develop better tools. Our future work is geared towards that direction.

**REFERENCES**


